



TECHNISCHE  
UNIVERSITÄT  
WIEN

# DMP-Unterstützung am Beispiel der TU Wien

Barbara Sánchez & Christiane Stork

Zentrum für Forschungsdatenmanagement

[barbara.sanchez@tuwien.ac.at](mailto:barbara.sanchez@tuwien.ac.at), [christiane.stork@tuwien.ac.at](mailto:christiane.stork@tuwien.ac.at)

[@RDMTUWien](#)

Wien, 29. Juni 2020



Diese Präsentation ist mit einer Creative Commons Attribution 4.0 International License (CC BY 4.0) lizenziert. Sánchez, Barbara, Stork, Christiane. Präsentation. DMP-Unterstützung am Beispiel der TU Wien. Juni 2020. Power Point Präsentation

Die Präsentation wurde für den Online-Workshop [Forschungsdatenmanagement in Österreich. Erfahrungsaustausch zu Beratung und Services](#) erstellt. Die Veranstaltung wurde gemeinschaftlich von FWF und TU Wien organisiert und im Rahmen des Projekts **FAIR Data Austria** ([forschungsdaten.at/projekte/fda](https://forschungsdaten.at/projekte/fda)) durchgeführt.



## Policy für FDM (seit Juli 2018)

- Umgang mit Forschungsdaten
- Nutzungsrechte, Aufbewahrung, Verantwortlichkeiten
- Erstellung und Aktualisierung von DMPs



## FWF Forschungsdatenpolicy (seit Jänner 2019)

- Open Data, DMP

## Horizon 2020 (seit 2014 Pilot, seit 2017 by default)

- Open Research Data, DMP

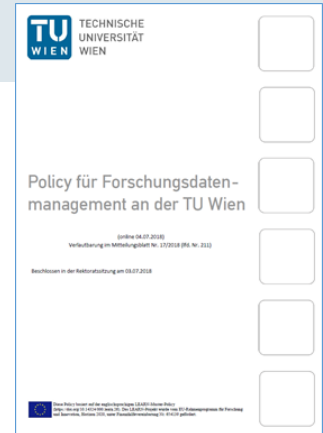


## Open data and Re-use of Public Sector Information (PSI)

- EU-Richtlinie, Überarbeitung im Juli 2019

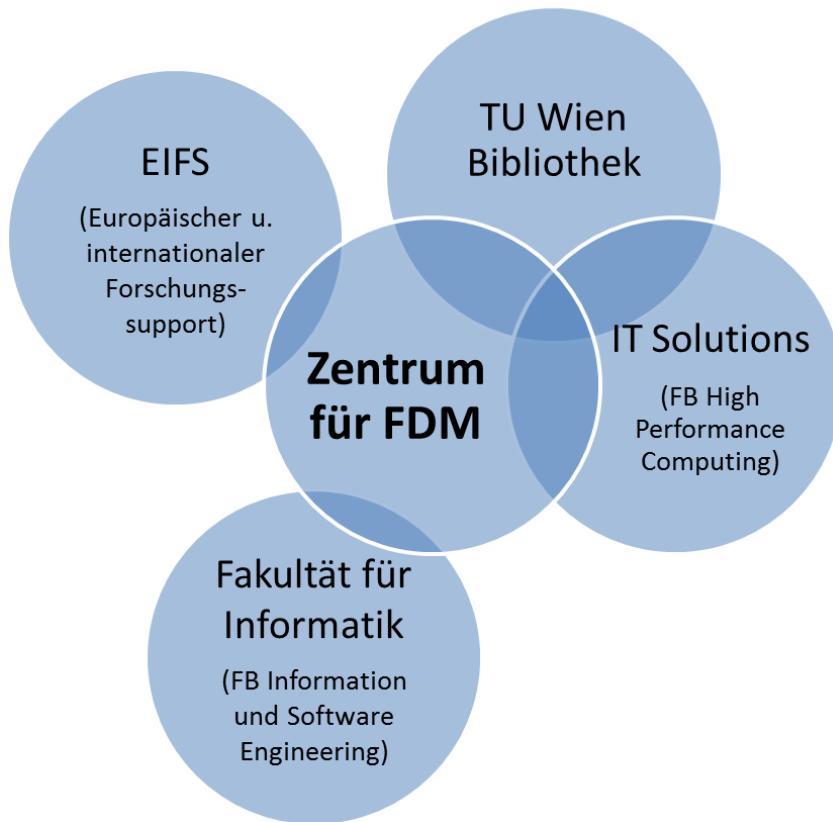
## European Open Science Cloud (EOSC)

- Harmonisierung von vorhandenen Infrastrukturen und Policies





# Das TU Wien FDM-Team



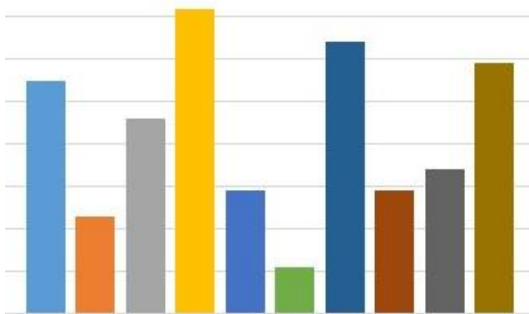
TU Wien FDM-Team bereichsübergreifend besetzt:

- 3 MA aus Zentrum für FDM
- 1 MA aus Forschungssupport (EIFS)
- 1 MA aus Fakultät für Informatik (Forschungsbereich Information und Software Engineering)

**Plus** Projektmitarbeiter\_innen aus BMBWF-finanziertem Projekt FAIR Data Austria:

- Software-Entwickler
- FAIR Data Officer

- **Wir erfassen**
  - Name, Institut, Fakultät
  - Kontaktweg (Wie wurden wir gefunden?)
  - Frage/Thema
  - Antwort

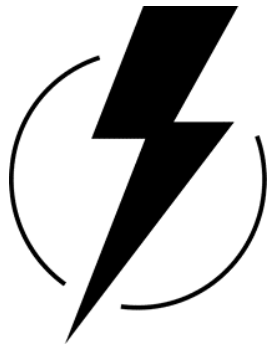


- **Nach ca. 1,5 Jahren**
  - 35 konkrete FDM-Anfragen
  - DMP (10)
  - Repositorien (9)
  - Weitere Themen
    - FDM-Policy, Software-Lizenzen, Informed Consent, IPR, Kosten, Registerforschung, Speicherung & Backup, TUprouCloud, Electronic Lab Notebook, Gitlab, FAIR
  - Top 5 Fakultäten bei Anfragen
    - Bauingenieurwesen, Elektrotechnik und Informationstechnik, Technische Chemie, Physik, Maschinenwesen und Betriebswissenschaften

## **FWF Forschungsdatenpolicy hat Impact**

- Beratungsszenarios
  - Erstbesprechung zum Thema DMP/FDM generell
  - Kommentierung eines vorhandenen DMP-Entwurfs
- Template wird als hilfreich empfunden
- Bei der ersten Version noch sehr unterschiedliche Auffassung von Begriffen
- Im Anfangsstadium viele Wiederholungen
- Je ausgereifter ein DMP, desto kürzer
- Verwirrung bei manchen Themenbereichen

- Auflistung von vielen unterschiedlichen Daten in Sektion 1 (*Description of the data*)



- Jedoch kein Hinweis auf diese Daten im weiteren Verlauf des Templates, z.B. bei den Themen Publikation und Lizenzierung

	<b>Data Officer</b>	<i>Who is responsible for the data management and the DMP of the project (name/email address)?</i>
<b>I</b>	<b>Data Characteristics</b>	
<b>I.1</b>	<b>Description of the data</b>	<i>What kinds of data/source code will be generated or reused (type, format, volume)? How will the research data be generated and which methods will be used? How will you structure the data and handle versioning? Who is the target audience?</i>
<b>II</b>	<b>Documentation and Metadata</b>	
<b>II.1</b>	<b>Metadata standards</b>	<i>What metadata standards (if any) will be in use and why? (see <a href="#">Digital Curation Centre</a>)</i>
<b>II.2</b>	<b>Documentation of data</b>	<i>What information is needed for the data to be findable, accessible, interoperable and re-usable (FAIR) in the future? Is the data machine-readable? How are you planning to document this information?</i>
<b>II.3</b>	<b>Data quality control</b>	<i>What quality assurance processes will you adopt? How will the consistency and quality of data collection be controlled and documented? (This may include processes such as repeat samples or measurements, standardised data capture, peer review of data or representation with controlled vocabularies.)</i>
<b>III</b>	<b>Data Availability and Storage</b>	
<b>III.1</b>	<b>Data sharing strategy</b>	<i>How and when will the data be shared and made accessible? What repository will you be using? What persistent identifier will be used?</i>
<b>III.2</b>	<b>Data storage strategy</b>	<i>What data are to be preserved for the long-term, and what data will not be stored? How and where will the data be stored and backed up during the research? How and where will the data be stored after the project ends? For how long will the data be stored? Are there any costs that need to be covered for storage? At what point during or after the project will the data be stored? Are there any technical barriers to making the research data fully or partially accessible?</i>





*„In meiner Disziplin gibt es keinen etablierten Metadatenstandard“*

*„Für meine Daten fallen keine Metadaten an“*

*„Was bedeutet machine-readable?“*

- Hinweis auf maschinengenerierte Metadaten
- Hinweis auf Datendokumentation (z.B. README Files, Codebooks, Infos aus Lab Journals, Algorithmen)
- Wichtig: **Dokumentation** ist als supplementary material im Repositorium zu **publizieren**
- Hinweis auf Metadatenfelder eines Repositoriums – dienen u.a. der Auffindbarkeit
- Machine-readable: z.B. eine Zahlentabelle als PDF nicht, in Form von CSV, JSON oder XLS schon
- FAIR: eher nicht als Begriff, jedoch als Konzept bekannt
- **FAIR ≠ Open**





# Tipp

Aufforderung an Forschende, die Perspektive eines Re-users einzunehmen und zu überlegen, welche Infos notwendig sind, um die Daten zu verstehen/zu zitieren/zu verwenden/zu verändern etc.

	<b>Data Officer</b>	<i>Who is responsible for the data management and the DMP of the project (name/email address)?</i>
<b>I</b>	<b>Data Characteristics</b>	
<b>I.1</b>	<b>Description of the data</b>	<i>What kinds of data/source code will be generated or reused (type, format, volume)? How will the research data be generated and which methods will be used? How will you structure the data and handle versioning? Who is the target audience?</i>
<b>II</b>	<b>Documentation and Metadata</b>	
<b>II.1</b>	<b>Metadata standards</b>	<i>What metadata standards (if any) will be in use and why? (see <a href="#">Digital Curation Centre</a>)</i>
<b>II.2</b>	<b>Documentation of data</b>	<i>What information is needed for the data to be findable, accessible, interoperable and re-usable (<a href="#">FAIR</a>) in the future? Is the data machine-readable? How are you planning to document this information?</i>
<b>II.3</b>	<b>Data quality control</b>	<i>What quality assurance processes will you adopt? How will the consistency and quality of data collection be controlled and documented? (This may include processes such as repeat samples or measurements, standardised data capture, peer review of data or representation with controlled vocabularies.)</i>
<b>III</b>	<b>Data Availability and Storage</b>	
<b>III.1</b>	<b>Data sharing strategy</b>	<i>How and when will the data be shared and made accessible? What repository will you be using? What persistent identifier will be used?</i>
<b>III.2</b>	<b>Data storage strategy</b>	<i>What data are to be preserved for the long-term, and what data will not be stored? How and where will the data be stored and backed up during the research? How and where will the data be stored after the project ends? For how long will the data be stored? Are there any costs that need to be covered for storage? At what point during or after the project will the data be stored? Are there any technical barriers to making the research data fully or partially accessible?</i>



- Unklarheiten bzgl. temporärer Speicherung und Langzeitspeicherung
- Häufig nicht klar, dass sich "Sharing" sowohl auf die Projektphase (Teilen mit Projektpartnern), als auch auf das Teilen (offen oder eingeschränkt) nach Projektende bezieht
- Verwechslung von OwnCloud und Repositorium
- Publikation von Daten: Welche? Alle?
- Aufbewahrung von Rohdaten?
- „Repositorium“ hat für Informatiker andere Bedeutung
- Wie erhält man einen PID?



## Themen in der Beratung

- Definition von „underlying data“
- Hinweis auf TUproCloud für temporäre Speicherung
  - tägliches Backup, Zugangskontrolle
  - auch für externe Projektpartner
- Erläuterung von Repositoriums-Funktionen
- Hinweis auf GitHub für Speicherung von Software/Code
- Aufbewahrungsdauer: 10 Jahre lt. TU Policy
- Hinweis auf Löschung

IV Legal and Ethical Aspects	
IV.1	<p><b>Legal aspects</b></p> <p><del>Are there any legal barriers to making the research data fully or partially accessible?</del>  <del>Who owns the data?</del>  <del>What licence for reuse are you planning to attach to the data?</del>  <del>Are there any restrictions on the re-use of the data? If so, why?</del></p>
IV.2	<p><b>Ethical aspects</b></p> <p>Are there any ethical barriers to making the research data fully or partially accessible?          If applicable, how are you planning to deal with sensitive data during and after the project?          Consider "<a href="#">Ethics for researchers</a>" published by the European Commission or "<a href="#">The European Code of Conduct for Research Integrity</a>".</p>
<p><b>No data will be generated or analysed</b> <i>The FWF recognises that some projects will not generate or analyse research data and similar materials. In these cases, a short explanation is required.</i></p>	



- Ownership? Rolle der Institution?
- Joint Ownership mit Projektpartnern?
- Ab wann Urheberrecht?
- Lizenzierung von Rohdaten/maschinengenerierten Daten? Creative Commons CC.0?
- Bedenken bei CC-By – Commercial Use
- Lizenzierung von Software?
- Personenbezogene Daten
- Überschneidungen zwischen den Bereichen legal & ethical (z.B. Informed Consent)

- Repositorien ermöglichen

- Speicherung/Archivierung
- Sharing
- Publikation
- Auffindbarkeit (F von FAIR)
- Zugang (A von FAIR)
- Vergabe von Persistent Identifier (z.B. DOI)
- Vergabe von Nutzungslizenz
- Zitation
- FAIRness

- Allgemeine Empfehlungen

- **Zenodo**: multidisziplinär, alle Datentypen, self-upload  
(<https://zenodo.org/>)



- **Github**: für Code/Software  
(<https://github.com/>)



- **AUSSDA**: quantitative Umfragedaten, Datachecks  
(<https://data.aussda.at/>)



- **re3data**: Register von 2.000 Datenrepositorien  
(<https://www.re3data.org/>)



- Suche nach trusted Repositorien (z.B. Core Trust Seal)

„Gibt es Mustertexte zu DMPs?“



„Kann ich DMP-Beispiele von anderen Forschenden sehen?“

Funding programme/ grant number	
Relevant policies and guidelines	<ul style="list-style-type: none"> <li>TU Wien Policy for Research Data Management: <a href="https://www.tuwien.at/fileadmin/Assets/dienstleister/Datenschutz_und_Dokumentationsmanagement/Policy_for_Research_Data_Management.pdf">https://www.tuwien.at/fileadmin/Assets/dienstleister/Datenschutz_und_Dokumentationsmanagement/Policy_for_Research_Data_Management.pdf</a></li> <li>Directives and Regulations of the TU Wien Rectorate: <a href="https://www.tuwien.at/en/tu-wien/organisation/service-providers/data-protection-and-document-management/directives-regulations/">https://www.tuwien.at/en/tu-wien/organisation/service-providers/data-protection-and-document-management/directives-regulations/</a></li> <li>Other (e.g. from your project partner)</li> </ul>

## I.1 Description of the data

### Good to know

Research data comes in all forms and sizes, from different sources and with varying potential for reuse. You can find a definition for research data in the [Policy for Research Data Management \(RDM\) at the TU Wien](#).

In the DMP you are asked to describe both: existing data you are reusing and data you are generating. You should list the types, formats and estimated volume of your data. Plus, in case of existing data, the source (data publisher) and any constraints on reuse.

### data types

- text documents (doc, odt, pdf, txt, etc.)
- graphics/images (jpeg, svp, png, gif, tiff, etc.)
- spreadsheets (xls, ods, csv, sas, stata, spss, etc.)
- structured text (html, json, tex, xml, etc.)
- video/film (mpeg, avi, wmv, mp4, etc.)
- databases (MS Access, MySQL, Oracle, etc.)
- software applications source code (c++, javascript, java, etc.)
- audio (mp3, wav, aiff, ogg, etc.)
- configuration data (ini, conf, etc.)
- etc.

### sources: generated data

- observations
- laboratory
- field instruments
- experiments
- simulations
- compilations
- sensor data
- observational data
- questionnaires
- etc.

### sources: existing data

- GIS data
- Statistik Austria
- data center
- repositories
- etc.

Online surveys:  
survey with the online survey tool XYZ.

### I.1 Description of the data (3)

How will you structure the data and handle versioning?

Filenames will follow the projects naming convention as defined in document ABC.

The filenames will include a timestamp of creation.

The respective work package leader will handle the structure and versioning of the research data. For the whole project a standardized folder structure and naming conventions will be in place.

### I.1 Description of the data (4)

Who is the target audience?

Target audience is the research community in ABC.

Target audience: decision makers at involved universities.

Target audience: researchers in environmental engineering and society as a whole.

### Guiding questions (FWF-template) and inspirations for your answers

#### I.1 Description of the data (1)

What kinds of data/source code will be generated or reused (type, format, volume)?

We will produce the following data:  
Excel files with lab data  
type: measurement data, format: XLS, volume: 100 MB

We will reuse the following data:  
questionnaire from XYZ, [https://doi.org/...](https://doi.org/)  
type: survey data, format: XLS, total volume: 10 MB

This project produces aggregated datasets in CSV format (10 x ~10 MB) that contain data points that combine alcohol consumption data with UFO sighting data and a correlation plot of these in PNG format (<10 MB).

The project reuses two external CSV datasets with statistical data:

- Alcohol Consumption: OECD (2018), Alcohol consumption (indicator). DOI: 10.1787/e6695909-en, 112 KB (accessed on 22 March 2018)
- Dataset: UFO Sightings, Sigmond, Axel (2014) UFO-reports <https://github.com/planetbeing/ufo-reports>, 13 MB (accessed on 22 March 2018)

Research outputs:  
- merged and pre-processed data generated from the initial collected data  
- Jupyter notebook including the CRISP-DM workflow and the final results of prediction and analysis

Files are stored in open, standardized formats as far as possible. The formats PDFa, CSV, MPEG-4 (audio track WAVE) and if necessary, TIFF are used for this purpose. Where conversion to an open format is not possible, original formats are saved.

We will develop models:  
based on literature review and prototype development  
type: dataset, format: OWL, volume 10 MB

#### I.1 Description of the data (2)

How will the research data be generated and which methods will be used?

Data will be produced in a laboratory by performing XYZ.

For this dataset, we are using the following instruments:  
Python (an interpreted, high-level, general-purpose programming language)  
Jupyter Notebook (an open-source web application that allows you to create and share documents that contain live code, equations, visualizations and explanatory text)

the raw data will be constantly accessible and attached within the same folder as the modified version. The e-data will be registered in folders following our standardized folder structure and naming convention as defined in our group.

### Example in tabular form

Source	Type	Format	Volume
questionnaires, reports	textual data	docx, pdf, txt	max. 500 MB
interview recordings	voice data	wav	5-10 GB
sensor data	structured data	csv, xlsx	1-2 GB
generated from existing data	ontologies	owl, xml	max. 1 MB

## II.1 Metadata standards

### Good to know:

Metadata are data about data. Metadata standards are useful to structure metadata in a commonly agreed way. You can find more information on [metadata and metadata standards](#) on our website.

The easiest way to provide clear metadata is to use discipline specific metadata standards. If you are not aware of any metadata standards used in your domain, you should check the following websites for help:

- [FAIRsharing](#)
- [re3data](#) (inter metadata standards)
- [Digital Curation Centre](#) (DCC)

If there is no suitable metadata standard for your data, you should use some kind of documentation (e.g. README file, see section II.2) to provide the relevant metadata. Make sure that data, concepts, columns, labels, etc. can be understood by people outside your research group - and outside your research domain as well.

Repositories provide common metadata templates that are domain independent. You should adhere to the repository metadata schema and provide as much information as possible when loading up data (title, description, keywords, type of publication, etc.).

If there is a metadata standard for your domain

- use it and declare it.
- find a repository that supports this standard, e.g. by using re3data
- if there is no repository, use the general purpose repository. In such a case you will have to provide common metadata such as title, description, etc. plus a README file as mentioned above.

# Ausbau von „Muster-DMPs“

- Stetige Erweiterung von Textmodulen
- Forschende motivieren zum
  - Sharing des DMPs mit Kolleg\_innen an der TU Wien
  - Hochladen des DMPs in die Projektdatenbank
  - Hochladen des DMPs in ein Repository (mit CC-By-Lizenz)
- Muster-DMPs für weitere Fördergeber





- „Projektdatenbank neu“
  - Pflichtfrage:  
*„Verlangt Ihr Fördergeber einen DMP für das Projekt?“*
  - Feld für Hochladen eines DMPs
- Machine-actionable DMP-Tool
  - Im Rahmen des Projekts FAIR Data Austria
  - Idee: automatischer Zugriff auf Informationen, die bereits an der Institution vorhanden sind (z.B. ORCID iD, Name, Institutszugehörigkeit, Projektinfos, etc.)
- Schnittstellen mit Systemen von Fördergebern
  - Umsetzung im Rahmen des Projekts RIS Synergy



# Workshops

- **Data management and research ethics**
  - für TU Wien Forschende
  - gemeinsam mit TU Wien Ethik-Koordinatorin & **Bibliothek**
- **Basics in proposal writing**
  - ZG PhD Studierende
- Kurzpräsentationen des Zentrums für FDM und Thema FDM bei diversen Veranstaltungen

# Kontakt und Information

✉ [research.data@tuwien.ac.at](mailto:research.data@tuwien.ac.at)

Webseite: [www.tuwien.at/researchdata](http://www.tuwien.at/researchdata)

Twitter: @RDMTUWien

The screenshot shows the TU Wien website interface. At the top, there are navigation links for 'TU WIEN', 'STUDIUM', and 'FORSCHUNG'. The main content area features a large image of a modern building interior with a person in a red coat walking. Overlaid on the image is a blue circular logo containing a database icon and binary code. Below the image, the text reads: 'Willkommen beim Zentrum für Forschungsdatenmanagement'. A sidebar on the left contains a menu with options like 'Übersicht', 'Forschungsdatenmanagem...', 'DMP', 'Förderrichtlinien', 'Speichern und teilen', 'Aufbewahren und publizieren', 'Persistente Identifikatoren', 'Team', 'Kontakt', and 'News'.

The screenshot shows the Twitter profile for 'TU Wien Research Data Management' (@RDMTUWien). The profile picture is a blue circular logo with a database icon and binary code. The bio states: 'We support researchers @tuvienna handling research data and code throughout the entire life cycle. #researchdata #RDMsupport #DMP #datasharing'. It also lists the location as Vienna, Austria, the website as [tuwien.at/en/researchdata](http://tuwien.at/en/researchdata), and the start date as July 2019. The profile has 192 followers and 277 following. A recent tweet from 3 hours ago mentions a workshop on June 29th with the funding agency FWF and research institutions regarding service offers.

# BMBWF-geförderte Digitalisierungsprojekte

# Digitalisierungs-Cluster „Forschungsdaten“

 Bundesministerium  
Bildung, Wissenschaft  
und Forschung

## Digitale und soziale Transformation

Ausgewählte Digitalisierungsvorhaben an öffentlichen Universitäten 2020 bis 2024

Open Science

Digitale Hochschulverwaltung

Soziale Dimension

Learning Analytics

Digitales Lehren und Lernen

1. RIS Synergy
  2. FAIR Data Austria
  3. Austrian DataLAB and Services
- Die Projekte fördern die **Zusammenarbeit zwischen österreichischen Universitäten** bei der Entwicklung kohärenter und solider Dienste für Forschungsdaten
  - Dadurch sichern sich österreichische Universitäten ihre **Rolle in der internationalen Forschungslandschaft**

Mehr Infos: <https://forschungsdaten.at/>

Download: siehe [hier](#)

# FAIR DATA AUSTRIA

Lead: **TU Graz**

Projektdauer: Jänner 2020 – Dezember 2021

Partner:

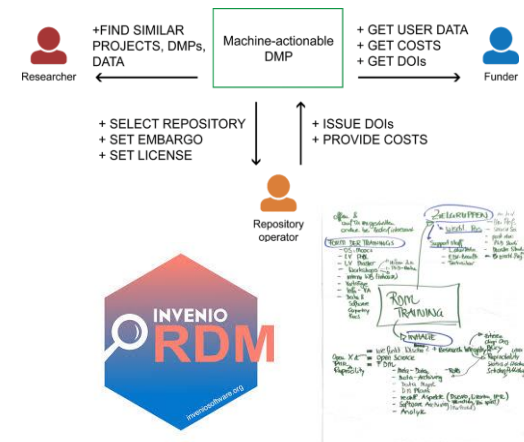


23 assoziierte Partner

“FAIR Data Austria trägt zur Stärkung des Wissenstransfers zwischen Universitäten, Wirtschaft und Gesellschaft bei und unterstützt die **nachhaltige Implementierung der European Open Science Cloud (EOSC) und der FAIR Prinzipien** („findable“, „accessible“, „interoperable“ und „re-usable“)

Implementierung der FAIR-Prinzipien durch

- ein integriertes Forschungsdatenmanagement (FDM)
- den Aufbau und Entwicklung von Next-Generation **Repositorien** für Forschungsdaten, Code und anderen Forschungsoutput
- die Entwicklung von **Training- und Support-Services**



[forschungsdaten.at/projekte/fda/](https://forschungsdaten.at/projekte/fda/)